

研究论文

基于递推部分最小二乘自适应质量监控策略 及其在橡胶混炼过程中的应用

宋 凯¹, 王海清², 李 平²

(¹ 天津大学化工学院, 天津 300072; ² 浙江大学工业控制技术研究所, 浙江 杭州 310027)

摘要: 提出一种新的基于递推部分最小二乘 (RPLS) 算法的自适应在线质量监控策略。利用隐变量选择算法, 根据实时采集的现场数据, 在不增加计算和存储容量的基础上递推更新 RPLS 过程监测模型, 进而更新 Q_c 控制限, 从而使 RPLS 自适应质量监控系统具有强时变跟踪特性, 能够有效克服传统监测算法 Q_c 无法反映系统时变性的缺点, 大大降低了监控系统的误报率和漏报率, 提高监控系统性能。并根据橡胶混炼过程特点, 将此方法运用于该时变间歇过程质量监控中, 取得了满意效果。

关键词: 递推部分最小二乘; 统计质量监控; 橡胶混炼

中图分类号: TQ 02; TP 29

文献标识码: A

文章编号: 0438-1157 (2007) 02-0410-07

RPLS based adaptive statistical quality monitoring of rubber mixing process

SONG Kai¹, WANG Haiqing², LI Ping²

(¹ School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China;

² Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, Zhejiang, China)

Abstract: An adaptive recursive partial least squares (RPLS) monitoring scheme is proposed to improve the time-variant tracking power of the statistical quality monitoring (SQM) systems. When new samples are obtained, this new method updates the monitoring model and the Q statistic control limit (Q_c) on the basis of the improved RPLS algorithms. Thus it could overcome the shortage of the traditional fixed SQM successfully. The theoretical findings were fully supported by the application performed on the rubber mixing process in a large-scale tire plant in east China. It was shown that the compounds quality is improved remarkably and the false alarm frequency was reduced significantly.

Key words: recursive partial least square; statistical quality monitoring; rubber mixing process

引 言

统计质量监控 (SQM) 方法能够有效利用历史数据对生产过程进行质量监控, 不需要严谨的过程机理模型, 甚至不需要对工艺过程的深入的先验知识, 因此是化工等复杂生产过程实现安全生产和

质量监控的主要手段^[1-3]。然而常规的 SQM 方法中, 模型一旦建立, 即存入模型库中, 不会随着过程特性的变化而在线更新, 不具有时变跟踪能力。而实际的工业过程 (尤其是间歇生产过程), 其操作参数和工况大多是时变的。催化剂的不断老化、设备磨损、传感器精度和零点的漂移, 以及定期的

2005-12-21 收到初稿, 2006-03-30 收到修改稿。

联系人: 王海清。第一作者: 宋凯 (1975—), 女, 博士, 讲师。

基金项目: 国家自然科学基金项目 (20576116)。

Received date: 2005-12-21.

Corresponding author: WANG Haiqing, associate professor.

E-mail: hqwang@iipc.zju.edu.cn

Foundation item: supported by the National Natural Science Foundation of China (20576116).

设备维修和清洗等,都会造成工业过程特性随着时间的缓慢变化。这些情况反映在过程模型中即可表现为^[4]: (1) 过程测量值均值的变化; (2) 过程测量值方差的变化; (3) 变量间相关结构的变化; (4) 模型隐变量个数的变化等。显然,用非时变的 SQM 方法对时变过程进行监测,不可避免地会发生误报或漏报现象,大大降低监控系统的可信性。如何有效利用不断累积的生产数据在线更新过程模型,使监控系统能够及时掌握生产过程的实时工况,具有一定的时变跟踪特性,及时杜绝安全隐患,实现对工业过程强有力的监控,成为目前的主要研究课题之一。

在各种统计方法中,部分最小二乘(PLS)方法能够通过提取互相正交的隐变量,在保留有效信息的情况下将高维变量空间压缩到维数大大降低的隐变量空间,克服测量信息的噪声、数据的缺失和变量间的强相关性^[5]。更为重要的是,与其他统计方法不同,PLS 没有忽略被控变量(质量变量)的信息,而是充分加以利用,建立更能反映系统关系的统计模型,因此近年来在化工过程的质量控制和在线监测等方面得到了广泛重视和研究^[6-7]。其扩展算法,递推部分最小二乘(RPLS)不仅具有传统 PLS 算法在克服变量间相关性和噪声干扰方面的优势,而且可以在线及时更新,有效跟踪系统的时变特性,计算量和存储容量不会随着样本量的增加而增加,在时变、间歇系统的建模和质量预测方面有着独特的优势^[8-9]。然而,常规 RPLS 算法要选取所有隐变量,导致此方法存在一定的过拟合现象,无法应用于故障诊断等质量监控方面。因此,本文对其进行改进,利用方差累计和(CPV)参数确定 RPLS 隐变量数,在更新监控模型的基础上,使得监控系统具有时变跟踪特性,能够根据新 RPLS 模型,及时调整 Q_a 控制限,克服传统监测方法中 Q 统计量无法跟踪系统变化而造成的误报或漏报现象,提高监控系统可靠性。

橡胶混炼过程是一个典型的快速间歇过程(仅几分钟),具有强烈的时变性,无可靠的机理模型,快检结果严重滞后(滞后时间是混炼周期的十几倍),数据噪声含量大,变量之间存在强相关性。这些特点使得无法应用传统的监控方法对其进行有效监控^[10]。因此,运用各种统计方法对混炼过程进行质量监控则成为实现橡胶混炼质量监控和安全生产中少数可行的研究策略之一。本文即以此为

例,来测试 RPLS 监控算法的时变跟踪性和有效性。

2 改进的 RPLS 自适应质量监控算法

为了有效利用新样本对模型进行更新,降低模型的计算和存储负担,Helland 等^[11]于 1992 年首先提出了 RPLS 算法。之后,Qin^[8]、宋凯等^[9]在此基础上做了修改和补充并将其应用于化工等复杂过程的建模与质量预测过程中。RPLS 算法不仅保留了传统 PLS 算法在克服变量相关性和抗噪声方面的独特优势,且能够在已有的 PLS 模型的基础上利用新数据对模型进行在线递推更新,计算量和数据存储量都不会随着样本的增加而增加。

2.1 常规 RPLS 算法和监测统计量

本节先简要回顾传统 RPLS 算法原理和常规监测统计量。

给定两个数据矩阵 $\mathbf{X} \in R^{m \times n}$ (过程变量)、 $\mathbf{Y} \in R^{m \times p}$ (质量变量),并且认为它们之间存在线性关系

$$\mathbf{Y} = \mathbf{X}\mathbf{C} + \mathbf{V} \quad (1)$$

式中 $\mathbf{V}$ 和 $\mathbf{C}$ 分别为相应维数的噪声和系数矩阵。

RPLS 回归算法通过对 $\mathbf{X}$ 和 $\mathbf{Y}$ 进行线性分解建立如下模型

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E}_k = \sum_{i=1}^k \mathbf{t}_i \mathbf{p}_i' + \mathbf{E}_k = \mathbf{T}\mathbf{P}' + \mathbf{E}_k \quad (2)$$

$$\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{F}_k = \sum_{i=1}^k \mathbf{u}_i \mathbf{q}_i' + \mathbf{F}_k = \mathbf{U}\mathbf{Q}' + \mathbf{F}_k \quad (3)$$

式中 $\hat{\mathbf{X}}$ 、 $\hat{\mathbf{Y}}$ 分别为矩阵 $\mathbf{X}$ 和 $\mathbf{Y}$ 的拟合矩阵; $\mathbf{E}_k$ 、 $\mathbf{F}_k$ 为相应的拟合误差矩阵; $\mathbf{t}_i$ 和 $\mathbf{u}_i$ 为第 i 个 RPLS 隐变量的得分向量; $\mathbf{p}_i$ 和 $\mathbf{q}_i$ 是相应的负载向量^[6,12]。在 RPLS 算法中,将得分向量(隐变量) $\mathbf{t}_k (k=1, 2, \dots, n)$ 标准化,而不是像经典 PLS 那样将轴向量 $\mathbf{w}_k$ 标准化。

引理 1^[8]: 如果 $\text{rank}(\mathbf{X}) = r \leq n$, 则

$$\mathbf{E}_r = \mathbf{E}_{r+1} = \dots = \mathbf{E}_n = \mathbf{0} \quad (4)$$

因此,当 RPLS 的隐变量数 k 满足 $r \leq k \leq n$ 时

$$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E}_k = \mathbf{T}\mathbf{P}' \quad (5)$$

由引理 1 可以导出 RPLS 更新算法: 当 $\mathbf{E}_k \leq \epsilon$ (ϵ 为任意小正数,代表允许的失配误差)时,给定两个 RPLS 子模型, $\{\mathbf{X}, \mathbf{Y}\} \xrightarrow{\text{RPLS}} \{\mathbf{T}, \mathbf{W}, \mathbf{P}, \mathbf{B}, \mathbf{Q}\}$ 和 $\{\mathbf{X}_1, \mathbf{Y}_1\} \xrightarrow{\text{RPLS}} \{\mathbf{T}_1, \mathbf{W}_1, \mathbf{P}_1, \mathbf{B}_1, \mathbf{Q}_1\}$, 对数据

对 $\begin{bmatrix} \mathbf{P}' \\ \mathbf{P}_1 \end{bmatrix}$, $\begin{bmatrix} \mathbf{BQ}' \\ \mathbf{B}_1 \mathbf{Q}'_1 \end{bmatrix}$ 进行 RPLS 建模得到的模型, 等价于对数据对 $\begin{bmatrix} \mathbf{X} \\ \mathbf{X}_1 \end{bmatrix}$, $\begin{bmatrix} \mathbf{Y} \\ \mathbf{Y}_1 \end{bmatrix}$ 进行 RPLS 建模所得到的模型。同理可以推广到有 s 组数据对的情况, 即对数据对

$$\{\mathbf{X}_i, \mathbf{Y}_i\} \xrightarrow{\text{RPLS}} \{\mathbf{T}_i, \mathbf{W}_i, \mathbf{P}_i, \mathbf{B}_i, \mathbf{Q}_i\} \quad i = 1, 2, \dots, s \quad (6)$$

进行 RPLS 建模, 同采用数据对 $\{\mathbf{X}_{\text{block}}, \mathbf{Y}_{\text{block}}\}$ 进行 RPLS 建模所得模型是等价的。其中

$$\mathbf{X}_{\text{block}} = \begin{bmatrix} \mathbf{P}'_1 \\ \mathbf{P}'_2 \\ \vdots \\ \mathbf{P}'_s \end{bmatrix}, \mathbf{Y}_{\text{block}} = \begin{bmatrix} \mathbf{B}_1 \mathbf{Q}'_1 \\ \mathbf{B}_2 \mathbf{Q}'_2 \\ \vdots \\ \mathbf{B}_s \mathbf{Q}'_s \end{bmatrix} \quad (7)$$

由上述分析可以看出, RPLS 算法不仅可以根据实时采集到的新数据进行在线更新, 具有良好的时变跟踪性能; 且计算量和数据存储容量由原来 $\mathbf{X}$ 的 $m \times n \times s$, 降低到 $\mathbf{P}$ 的 $n \times k \times s$ 。当样本数远远大于变量个数, 即 $m \gg n$ 时, RPLS 算法可以节省大量的计算时间和数据存储容量; 样本数越多, RPLS 算法的这一优势越明显。实际的工程应用中多为样本数远远大于变量数的情况。例如, 在橡胶混炼过程中, 过程变量通常只有温度、压力、能量等几个参数, 然而随着时间的累积, 样本却可以有成千上万个。

常规 SQM 监控方法主要通过 Q (SPE) 统计量对生产过程进行故障诊断和质量监控。正常工况下, 样本的 Q 统计量会在一定统计范围内波动, Q 统计量定义为

$$Q = \tilde{x}\tilde{x}' = \|\tilde{x}\|^2 \leq Q_\alpha \quad (8)$$

式中 $\tilde{x}$ 为未建模部分, 服从 χ^2 分布^[13]。由于 Q 统计量监测的是系统的未建模部分, 反映的是过程的物料平衡、能量平衡和正常的操作限制。当 Q 统计量发生变化而超限时, 说明当前统计模型所代表的正常工况下的过程关系被破坏, 系统发生过程故障或传感器故障。 Q_α 为 Q 统计量的 α 置信限, 根据模型的统计特性计算得到。 Q_α 值确定得是否合理, 反映了系统模型是否能够正确揭示当前过程关系。 Q_α 值过高, 微弱幅值的故障将无法被检测, 系统出现漏报现象; 相反, Q_α 值过低, 正常扰动或者噪声干扰也可能引起样本 Q 统计量超限, 系统过于敏感而出现误报。由此可见, 为了提高监控系统的可靠性, Q_α 控制限必须能够准确反映系统的当前工况, 具备时变跟踪特性。然而常规监测方

法中, 模型一旦确立, Q_α 随之确定并存入模型库中, 只有当重新更新系统模型库的时候才可能得到更新, 导致监控系统频繁的误报或漏报现象。

2.2 改进的 RPLS 质量监控算法

由前述对 RPLS 算法的分析可知, 虽然 RPLS 算法可以在线更新, 具有很强的时变跟踪能力, 且计算量不随样本数的增加而增大。但 RPLS 算法成立的条件是 $E_k \leq \epsilon$, 即认为过程关系完全可以被 RPLS 模型拟合, 不存在未建模部分, 即未建模方差 (或样本 Q 值) 几乎为零。这样不仅会由于隐变量选择过多而造成过拟合现象, 且无法进行统计质量监控。因此本文利用恰当的隐变量选择算法对其进行改进, 进而在此模型基础上获得可递推更新的 Q_α 控制限, 实现对时变过程的质量监控。

目前存在着很多隐变量 (LV) 优化选择算法^[4,14-15]:

- autocorrelation;
- cross-validation;
- cumulative percent variance (CPV);
- scree test;
- average eigenvalues;
- imbedded error function;
- Xu and Kailath's approach;
- Akaike information criterion;
- minimum description length criterion;
- variance of reconstruction error.

上述算法虽然都可以用来优化隐变量个数, 但并不都适用于 RPLS 算法和在线质量监控。考虑到实用性和计算复杂度, 本文选用 CPV 参数来优化 RPLS 算法隐变量的选择。利用方差累计和挑选隐变量数, 在线更新 Q_α 控制限, 使得监控算法具有时变跟踪特性, 能够随过程的变化而变化。

方差累计和 (CPV) 定义为

$$\text{CPV}(k) = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^n \lambda_j} \times 100\% \quad (9)$$

式中 λ_j 为 $\mathbf{X}$ 的相关系数矩阵 $\mathbf{R} = \frac{1}{m-1} \mathbf{X}' \mathbf{X}$ 的特征值。令 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$, 当方差累计和大于某一给定值 (如 85%) 时, 相应的 k 即为所需 RPLS 模型的隐变量数。此时未建模方差可近似为

$$\tilde{\mathbf{A}} = \text{diag}(\lambda_{k+1}, \lambda_{k+2}, \dots, \lambda_n) \quad \lambda_{k+1} \geq \lambda_{k+2} \geq \dots \geq \lambda_n \geq 0 \quad (10)$$

常规 Q 统计量的控制限可由式 (11) 计算^[12]

$$Q_\alpha = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 + 1)}{\theta_1^2} \right]^{1/h_0} \quad (11)$$

其中

$$\theta_i = \sum_{j=k+1}^n (\lambda_j)^i, \quad h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2}$$

c_α 为正态分布的 $(1-\alpha)$ 分位线。

由引理 1 可知 RPLS 算法中, $\mathbf{T}$ 已被正交化, 因此

$$\mathbf{X}'\mathbf{X} = \mathbf{P}\mathbf{T}'\mathbf{T}\mathbf{P}' = \mathbf{P}\mathbf{P}' \quad (12)$$

当得到一组新的数据对 $\{\mathbf{X}_1, \mathbf{Y}_1\}$ 后,

$$\mathbf{X}_{\text{new}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}_1 \end{bmatrix}$$

$$\mathbf{X}'_{\text{new}} \mathbf{X}_{\text{new}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}_1 \end{bmatrix}' \begin{bmatrix} \mathbf{X} \\ \mathbf{X}_1 \end{bmatrix} = \mathbf{X}'\mathbf{X} + \mathbf{X}'_1 \mathbf{X}_1 =$$

$$\mathbf{P}\mathbf{P}' + \mathbf{P}_1 \mathbf{P}'_1 = \begin{bmatrix} \mathbf{P}' \\ \mathbf{P}'_1 \end{bmatrix}' \begin{bmatrix} \mathbf{P}' \\ \mathbf{P}'_1 \end{bmatrix} = \mathbf{P}_{\text{new}} \mathbf{P}'_{\text{new}} \quad (13)$$

因此 $\mathbf{X}_{\text{new}}$ 相关系数矩阵 $\mathbf{R}_{\text{new}}$ 的特征值 $\lambda_{j\text{new}}$ 可以通过 $\mathbf{P}_{\text{new}}$ 来求取。即

$$\mathbf{R}_{\text{new}} = \frac{1}{m_{\text{new}} - 1} \mathbf{X}_{\text{new}} \mathbf{X}'_{\text{new}} = \frac{1}{m_{\text{new}} - 1} \mathbf{P}_{\text{new}} \mathbf{P}'_{\text{new}} \quad (14)$$

由式 (9) 和式 (13) 可知

$$\begin{cases} \theta_{1\text{new}} = \sum_{j=k+1}^n \lambda_j = \text{trace}(\mathbf{R}_{\text{new}}) - \sum_{j=1}^k \lambda_j \\ \theta_{2\text{new}} = \sum_{j=k+1}^n \lambda_j^2 = \text{trace}(\mathbf{R}_{\text{new}}^2) - \sum_{j=1}^k \lambda_j^2 \\ \theta_{3\text{new}} = \sum_{j=k+1}^n \lambda_j^3 = \text{trace}(\mathbf{R}_{\text{new}}^3) - \sum_{j=1}^k \lambda_j^3 \end{cases} \quad (15)$$

式中 k 为利用方差累计和 CPV 所确定的隐变量个数。由式 (11) 可计算 $Q_{\alpha, \text{new}}$ 。

由此可见, 利用 RPLS 算法和 CPV 指标相结合, 不仅可以有效利用新数据对模型进行及时在线更新, 且计算量大大减少 $[\dim(\mathbf{P}_{\text{new}}) \ll \dim(\mathbf{X}_{\text{new}})]$ 。

注意: 因为 RPLS 算法是利用过程的稳态正常数据进行建模的, 所以当检测到一组新的数据组后, 要首先计算其 Q 统计量, 判断这组数据中是否含有故障数据或不可靠数据。一旦发现故障数据, 应及时报警, 提醒工艺人员注意, 同时将这组数据摒弃。只有当数据组中不含有不可靠数据时, 这组新的数据才可用于更新 RPLS 模型, 进而更新 Q_α 控制限。

基于 RPLS 的质量监控算法可归纳为:

(1) 对数据组 $\{\mathbf{X}, \mathbf{Y}\}$ 归一化 (均值为零, 方差为 1);

(2) 建立 PLS 模型

$$\{\mathbf{X}, \mathbf{Y}\} \xrightarrow{\text{PLS}} \{\mathbf{T}, \mathbf{W}, \mathbf{P}, \mathbf{B}, \mathbf{Q}\};$$

(3) 直到 $\|\mathbf{E}_k\| \leq \epsilon$, ϵ 为任意小正数;

(4) 用 CPV 参数优化隐变量数;

(5) 按式 (11) 计算 Q_α ;

(6) 当获得一组新的数据 $\{\mathbf{X}_1, \mathbf{Y}_1\}$ 时, 按步骤

(1) 的方法对其进行归一化, 建立新的 RPLS 子模型

$$\{\mathbf{X}_1, \mathbf{Y}_1\} \xrightarrow{\text{RPLS}} \{\mathbf{T}_1, \mathbf{W}_1, \mathbf{P}_1, \mathbf{B}_1, \mathbf{Q}_1\};$$

(7) $\mathbf{P}_{\text{new}} = [\mathbf{P} \mathbf{P}_1]$, $m_{\text{new}} = m + m_1$, 根据式

(9) CPV 指标确定隐变量个数;

(8) 由式 (15)、式 (11) 计算 $Q_{\alpha\text{new}}$;

(9) 计算 Q_{new} , 判断是否有故障数据, 如果有, 则报警, 并放弃此样本组, 返回步骤 (2), 否则, 转入步骤 (7);

(10) 令 $\mathbf{X} = \begin{bmatrix} \mathbf{P}' \\ \mathbf{P}'_1 \end{bmatrix}$, $\mathbf{Y} = \begin{bmatrix} \mathbf{B}\mathbf{Q}' \\ \mathbf{B}_1\mathbf{Q}'_1 \end{bmatrix}$, 返回步骤 (2)。

3 面向橡胶混炼的 RPLS 自适应质量监控系统

3.1 密炼机橡胶混炼过程

密炼机橡胶混炼过程是指在密炼机中将各种配合剂混入生胶中制成质量均匀的混炼胶的生产过程^[10,16]。

混炼过程是橡胶加工工艺中最基本和最重要的工艺过程之一。对混炼胶的质量要求主要有两个方面: 一是胶料的质量要保证橡胶制品具有良好的物理力学性能; 二是胶料本身要具有良好的工艺加工性能。为此, 在混炼过程中必须使各种配合剂完全而均匀地分散于生胶中, 保证胶料的性能均匀一致; 还应使胶料具有一定的门尼黏度以保证后续各项加工操作能顺利进行。

由于混炼胶的门尼黏度直接影响到后续加工的顺利进行和终端产品的质量, 因此对橡胶混炼过程的质量监控是橡胶制品质量控制的关键。如前所述, 橡胶混炼是一个典型的快速间歇过程, 具有强烈的时变特性。虽然随着生产的进行会不断累积大量的过程数据, 但由于常规监控方法的限制, 实时数据无法被有效利用, 模型和 Q 统计量的监测控制限不会在线更新, 因此不可避免地发生漏报和误报现象, 大大降低了监控系统的有效性。因此, 如

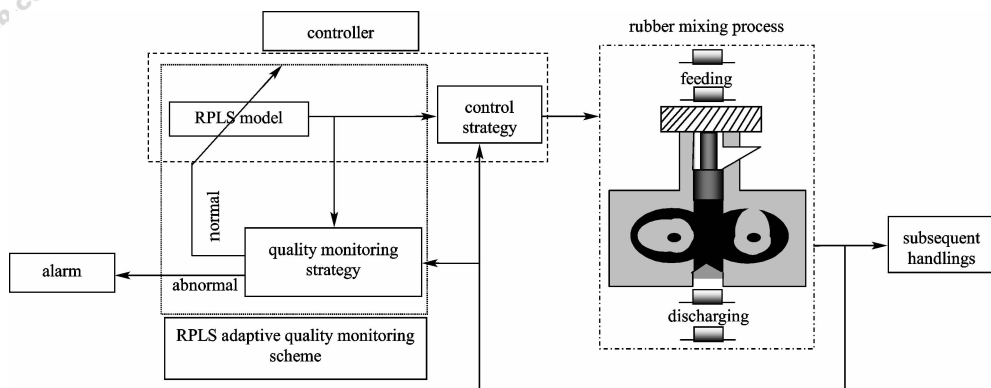


图 1 橡胶混炼过程 RPLS 自适应质量监控系统

Fig. 1 RPLS adaptive quality monitoring for rubber mixing process

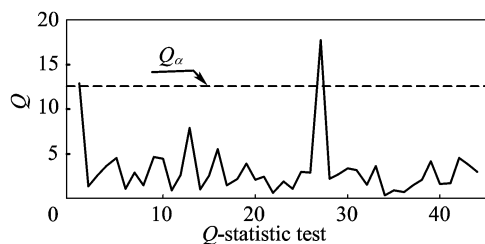
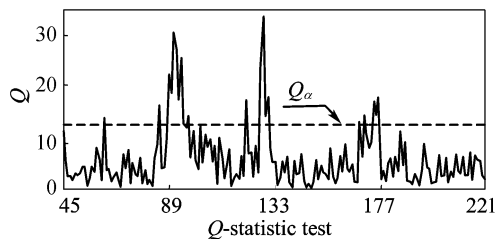
何有效利用生产数据, 建立具有时变更新能力的监控系统具有重要的理论和经济意义。

橡胶混炼质量监控系统流程如图 1 所示。监控系统在已有的模型基础上对当前胶料质量进行预测, 根据工艺要求制定控制策略 (如排胶、继续混炼等); 同时对新采集到的样本进行质量监测, 如果发现故障样本, 则发出报警信号并摒弃当前数据, 如未发现故障样本, 则利用新数据对过程监控模型进行更新。

橡胶混炼过程原始数据噪声含量大, 变量之间具有强相关性, 且影响混炼胶门尼黏度的因素很多, 其中原胶的质量 (主要是初始门尼黏度的均一性) 是最主要因素。本文以混炼胶门尼黏度为主要的被监控质量变量, 所用数据为国内某大型橡胶厂 2# 密炼机上采集的现场数据。为了验证 RPLS 监测算法的性能, 选用间隔时间较长、配方波动比较大的几组数据作为训练和测试样本, 并将其结果与常规 SQM 算法 (以 PLS 算法为例) 结果进行对比。所用样本分别为 2001 年 11 月 13、21、22、27 日的正常工况生产数据共 220 组, 每 44 组样本组成一个数据组。根据生产经验及 RPLS 算法建模结果, 控制变量确定为: x_1 —排胶点功率; x_2 —排胶点能量; x_3 —排胶点时间; x_4 —加油点功率; x_5 —加油点能量; x_6 —加油点时间; x_7 —投炭黑点能量; x_8 —投炭黑温度; x_9 —混炼初始温度。

3.2 普通 PLS 模型质量监测结果

以第一批 44 组生产数据为建模样本, 建立普通 PLS 模型, 并计算相应的 Q 统计量控制限 ($\alpha = 0.99$)。用此模型对其后的 156 组样本进行监控, 常规 Q 统计量监控图如图 2、图 3 所示。

图 2 建模样本常规 PLS 模型 Q 统计量监控图Fig. 2 Q -statistic test of first modelling samples of traditional PLS model图 3 检验样本常规 PLS 模型 Q 统计量监控图Fig. 3 Q -statistic test of traditional PLS model

86 个样本的 Q 统计量值均已明显超过控制限, 且超过 Q 统计量值控制限的样本已经连成一片。其后第 114 到第 132、第 172 到第 177 个样本的 Q 统计量值也已经超过控制限, 说明过程中已经发生故障。然而如前所述, 所选用的样本均为正常工况下采集的数据, 过程中无故障发生。监控系统发生了明显的误报现象。由此可见, 普通 PLS 算法由于模型一旦建立就不再更改, 因此无法跟踪过程时变特性。

3.3 自适应 RPLS 模型质量监控结果

在自适应 RPLS 监控算法中, 本例的 CPV 控制限定为 85%, 即方差累计和大于或等于 85% 时, 相应的 k 即为所需 RPLS 模型的隐变量数。相应于普通 PLS 监控算法, 训练和测试样本的质量监控

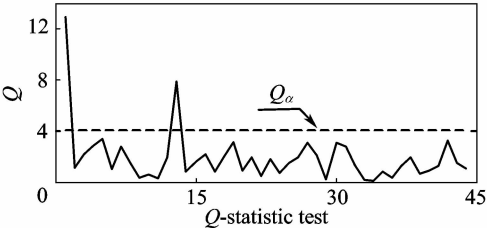


图4 建模样本 RPLS 模型 Q 统计量监控图

Fig. 4 Q-statistic test of first modelling samples of RPLS model

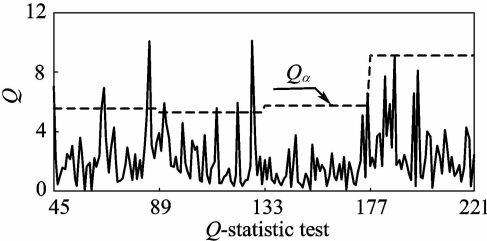


图5 检验样本 RPLS 模型 Q 统计量监控图

Fig. 5 Q-statistic test of RPLS model

图如图4、图5所示。

从图2和图4的结果可以看出，虽然两种监控算法中训练样本的Q统计量值以及相应的 Q_α 控制限的取值不同，但所揭示的监控信息互相吻合，即训练样本所反映的生产工况处于正常状态。

图5 RPLS 模型的质量监控结果中，只有不连续的8个样本点的Q统计量值超过控制限（测试样本共176个），其余样本均处于正常范围内。从统计学角度考虑，零星的几个不连续样本点的Q统计量值超限，通常是由于过程中不可测的工业噪声等随机现象引起的，过程仍然处于正常的生产工况，无故障发生。这与前述样本采集的实际情况完全相符。更为重要的是，图5中，控制限 Q_α 并不是固定不变的，而是随着过程的实际工况而相应地波动。这表明RPLS模型可以利用实时生产数据对模型进行及时更新，能够更好地跟踪过程的时变特性，降低常规算法非时变缺陷所导致的误报率和漏报率，大大提高监控系统的可靠性和有效性。

有关PLS和RPLS监控模型的参数细节如表1所示。

利用本文所述方法建立的质量监控系统在国内某大型橡胶厂2[#]密炼机上进行了测试，通过该厂几个月的实际炼胶生产的快检记录表明，监控系统的误报率和漏报率均大大降低。有助于工艺人员准确了解混炼过程实际状况，采取正确控制措施，极

大地降低了混炼胶质量的波动，减少了对胶料的重复混炼，从而大大提高了经济效益。

表1 PLS 和 RPLS 监控模型参数对比

Table 1 Details of PLS and RPLS monitoring model

Item	Algorithm					
	PLS			RPLS		
number of data set	1—5	1	2	3	4	5
number of LVs (<i>k</i>)	2	4	4	5	5	4
Q_α	14.98	4.04	5.540	5.293	5.744	9.121

4 结 论

本文通过CPV参数与RPLS算法相结合，建立可以在线更新的过程质量监控算法，计算未拟合方差。克服RPLS算法因不考虑未拟合误差（认为过程能够完全被模型拟合）而无法应用于质量监控的不足。并在此基础上对Q统计量控制限进行在线更新，使得控制限 Q_α 能够及时反映过程工况的波动，具有时变跟踪能力。并将此方法用于橡胶混炼过程的质量监控。结果表明，此方法使得质量监控系统能够更好地跟踪系统的时变特性，大大减少误报率和漏报率，可靠性得到明显提高。

此外如何有效利用“折息因子”、“遗忘因子”等方法进一步提高算法的自适应跟踪能力，扩展算法的灵活性和适用范围具有很强的理论和实用价值，有待于进一步深入研究。

符 号 说 明

- B ——得分向量拟合系数矩阵
- C ——拟合系数矩阵
- CPV——方差累计和
- c_α ——正态分布的 $(1-\alpha)$ 分位线
- E, F ——分别为过程、质量变量拟合误差矩阵
- k ——隐变量维数
- m ——样本维数
- n ——过程变量维数
- P ——对应于 X 的负载矩阵
- p ——质量变量维数
- p, Q ——分别为对应于 X 的负载向量、矩阵
- Q ——监测统计量
- Q_α ——监测统计量控制限
- q ——对应于 Y 的负载向量
- R —— X 的相关系数矩阵
- R ——实数空间

r ——矩阵 X 的秩

s ——数据组个数

T ——对应于 X 的得分矩阵

t ——对应于 X 的得分向量

u ——对应于 Y 的得分向量

V ——噪声矩阵

W ——对应于 X 的轴矩阵

w ——对应于 X 的轴向量

X ——过程变量

$\hat{X}$ ——过程变量拟合矩阵

x —— X 的某一行

$\tilde{x}$ —— x 未建模部分

Y ——质量变量

$\hat{Y}$ ——质量变量拟合矩阵

α ——置信度

ϵ ——任意小正数

θ_i ——未建模特征值 i 次方和

Λ —— R 的特征值组成的对角矩阵

$\tilde{\Lambda}$ ——未建模方差对角矩阵

λ —— R 的特征值

References

- [1] Venkatasubramanian Venkat, Rengaswamy Raghunathan, Kavuri Surya N, Yin Kewen. A review of process fault detection and diagnosis (I): Quantitative model-based methods. *Computers & Chemical Engineering*, 2003, **27**: 293-311
- [2] Venkatasubramanian Venkat, Rengaswamy Raghunathan, Kavuri Surya N, Yin Kewen. A review of process fault detection and diagnosis (II): Qualitative models and search strategies. *Computers & Chemical Engineering*, 2003, **27**: 313-326
- [3] Venkatasubramanian Venkat, Rengaswamy Raghunathan, Kavuri Surya N, Yin Kewen. A review of process fault detection and diagnosis (III): Process history based methods. *Computers & Chemical Engineering*, 2003, **27**: 327-346
- [4] Li Weihua, Yue H Henry, Sergio Valle-Cervantes, Qin S Joe. Recursive PCA for adaptive process monitoring. *Journal of Process Control*, 2000, **10**: 471-486
- [5] Hoskuldsson A. PLS regression methods. *J. Chemometrics*, 1988, **2**: 211-228
- [6] Wise B M, Gallagher N B. The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control*, 1996, **6** (6): 329-348
- [7] Kruger U, Chen Q, Sandoz D J, Mcfarlane R C. Extended PLS approach for enhanced condition monitoring of industrial processes. *AIChE Journal*, 2001, **47** (9): 2076-2091
- [8] Qin S J. Recursive PLS algorithms for adaptive data modeling. *Computers & Chemical Engineering*, 1998, **22**: 503-514
- [9] Song Kai (宋凯), Wang Haiqing (王海清), Li Ping (李平). Discounted-measurement RPLS algorithm and its application to quality control of the rubber mixing process. *Journal of Chemical Industry and Engineering (化工学报)*, 2004, **6**: 942-946
- [10] Chen Yaoting (陈耀庭). Techniques of the Rubber Products (橡胶加工工艺). Beijing: Chemical Industry Press, 1982
- [11] Helland K, Berntsen H E, Borgen O S, Martens H. Recursive algorithm for partial least squares regression. *Chemometrics Intell. Lab. Syst.*, 1992, **14**: 129-137
- [12] Svante Wold, Michael Sjostrom, Lennart Sriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 2001, **58**: 109-130
- [13] Chiang Leo H, Braatz Richard D, Russell Evan. Fault Detection and Diagnosis in Industrial Systems. London, New York: Springer, 2001
- [14] Wold S. Cross validatory estimation of the number of components in factor and principal component analysis. *Technometrics*, 1987, **20**: 397-406
- [15] Lindgren Fredrik, Geladi Paul, Rannar Stefan, Wold Svante. Interactive variable selection (IVS) for PLS (I): Theory and algorithms. *Journal of Chemometrics*, 1994, **8**: 349-363
- [16] Zheng Xiufang (郑秀芳), Zhao Jiashu (赵嘉澍). Equipments of the Rubber Plants (橡胶工厂设备). Beijing: Chemical Industry Press, 1984